UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/714,724	11/16/2000	Franck Barilloud	AUS9-2000-0483-US1	9479

7590 02/18/2005

Joseph R Burwell
P. O. Box 28022
Austin, TX 78755-8022

EXAMINER	
KIANERSI, MITRA	

ART UNIT	PAPER NUMBER
2145	

DATE MAILED: 02/18/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/714,724	FRANCK BARILLAND	
	Examiner	Art Unit	
	Mitra Kianersi	2145	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 03 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on 15 September 2004.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) 1-36 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) 1-36 is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on 16 November 2000 is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application (PTO-152)

6) ☐ Other: _____.

Response to Arguments

Applicant's argument filed on Sep/15/2004 has been fully considered, but they are not persuasive.

Applicant on page 4, lines 9 argues that the rejection has failed to address the first element of claim 1, unresponsive to a request from a requesting client for a distributed service, forwarding the request to a first distributed service manager associated with the requesting client. In addition, the rejection has also failed to address the fourth element of claim 1, "if the first distributed service manager does not have information about the distributed service, retrieving information about the distributed service from a second distributed service manager a: caching the retrieved information at the first distributed service manager". Gher et al on col 3, lines 5-25) discloses that the architecture enables service overloads and outages to be automatically and immediately addressed and later cleared without undo system reconfiguration overhead. The dynamic server switching system does not have to diagnose failures but simply responds to a failure of client-server communication. Also, Gher et al. on col 4, lines 41-58 disclose that in client-server system operations, multiple servers are provided to perform a particular function for the overall system, such as access to data stored in an automated cartridge library system, to ensure both that the plurality of clients receive an adequate level of service and a high level of server availability is maintained. Each client directs requests to a designated primary server and multiple clients typically access each server. The fault tolerance aspect of the system architecture of the dynamic server switching system makes use of a client communication interface based control which enables a client to simply redirect requests from an unresponsive server to a predetermined alternate server without the overhead of prior art systems. This is accomplished by maintaining data in the client communication interface Ci located in each client, which data identifies the primary server for that client and the preferred communication method as well as a hierarchy of successive alternate servers and communication method pairs. Gher et al. in claim 1 disclose means, responsive to a one of said plurality of clients failing to receive service

from said primary server defined for said one client, for retrieving an entry from said stored routing data to identify an available one of said least one alternate server designated for said one of said plurality of clients; and means, responsive to said retrieved entry, for directing service requests that are received from said one of said plurality of clients to said identified one of said at least one alternate server as defined by said retrieved data.

Applicant on page 4, line 14, argues that in the event that the client does not have requests served by the designated primary server or the designated communication method, the client communication interface traverses the list to ascertain the identity of the first available alternative server-communication method pair. The client then uses this retrieved data to initiate future requests. When an alternate server is being used, the client periodically tests the primary server-communication method pair to determine whether the fault has been cleared. If so, the client reestablishes the originally elected primary server-communication method pair as the request route, while wrapping up the existing communications with the alternate server-communication method pair. Gher et al. also in claim 1, discloses that storing routing data that defines a hierarchical ordering of servers for each of said plurality of clients, including a primary server and at least one alternate server designated for each of said plurality of clients, in a memory; retrieving, in response to a one of said plurality of clients failing to receive service from said primary server defined for said one client, an entry from said stored routing data located in said memory to identify an available one of said least one alternate server designated for said one of said plurality of clients; and directing, in response to said retrieved entry, service requests that are received from said one of said plurality of clients to said identified one of said at least one alternate server as defined by said retrieved data. Gher et al. on col 11, lines 24-41 disclose that the dynamic server switching system maintains a list in each client communications interface component which identifies the primary server for that client and the preferred communication mode as well as a hierarchy of successively secondary servers and communication method pairs. In the event that the client does not have requests served by the designated primary server or the designated communication method, then traverses

the list to ascertain the identity of the first available alternative server-communication path pair. The client then uses this retrieved data to initiate future requests. The client periodically tests the primary server-communication method pair to determine whether the fault has been cleared. If so, the client reestablishes the originally selected primary server-communication method pair as the request route. Since each client has its own list, the load presented to each of the servers can be balanced across the available servers.

Applicant on page 7, line 13 argues that not only does Gehr et al. do not disclose any features that are similar or analogous to the fourth element of claim 1, but Gehr et al. also does not disclose a series of steps that are similar or analogous to those performed at a distributed service manager or at some entity distinct from a client. Moreover, Applicant asserts that the steps that are recited in claim l are not performed at a client, but the rejection of claim l improperly interprets the elements of claim 1 by referring to certain steps in Gehr et al. that are performed by the client. Although it is not explicitly stated with respect to claim 1, the argument in the rejection of claim 1 seems to be based on a supposed equivalency between a client and a server; for example, claim 4 explicitly states that Gehr et al. discloses "that a server S1-S4 can be self-serving in that it can also function as a client for some service requests". Assuming arguendo that a server performs steps that are similar to the second element and third element of claim 1, this is irrelevant with respect to the claim as a whole; the claim specifically recites that the request originates at a client and that the information is returned to the requesting client. Gehr et al, argues that if there is a fault condition present in the data processing system, in the form of a failure to deliver the message to the server S1, or a failure of the server S1 to process the request, or the response message not returning to the client C1, or the server or communication methods being overloaded and messages being delayed or blocked, the timeout timer expires at step 514 to indicate a system fault or an exception message is returned to the client process C1 to indicate an error. Col 6, lines 59-67).

Applicant on page 8, line 14, argues that the Office action has presented a common rejection for independent claims 3, 15 and 25 because they have corresponding elements. It should be noted that if the applicant's argument are applicable to claims 15 and 27 and their dependent claims is because of their corresponding elements and

Applicant on page 9, line 10, argues that the rejection fails to explain what entities in Gehr et al. fulfill the roles of the distributed service manager and the local service manager in the present invention. Applicant asserts that it is not possible to present a proper anticipation argument based on Gehr et al. because Gehr et al. clearly does not disclose analogous or equivalent features to the present invention. Gher et al. on claim 1, disclose means for storing routing data that defines a hierarchical ordering of servers for each of said plurality of clients, including a primary server and at least one alternate server designated for each of said plurality of clients.

Applicant on page 9 line 22 argues that dependent claim 7 recites a feature in which a a distributed service manager broadcasts a request for a networked service to all distributed service managers if the distributed service manager does not have the necessary information. Gehr et al. does not disclose a broadcast operation. In fact, the rejection of claim 7 merely repeats the same argument that was applied against claim 1 without attempting to explain how Gehr et al. discloses any features that are analogous or equivalent to the broadcast operation. Gher et al. on col 1, lines 50-55 disclose that the server group must identify the failed server and then manage the shutdown of the failed server and its replacement. This fault recovery process is processing resource intensive. Furthermore, server group state changes must be broadcast to the members of the server group, which requires close timing synchronization in the presence of non-uniform communication services, in order to preserve a consistent state among the servers located in the group.

Applicant on page 11, line 21 argues that Gehr et al. and Kadansky et al. clearly fail to disclose or to suggest at least one feature of the present invention as recited within each independent claim, notwithstanding the argument: rendering Gehr et al. and

Kadansky et al. incapable of being used as primary and secondary references as argued by the current rejection. Moreover, a hypothetical combination of Gehr et al. and Kadansky et al. would also fail to reach the claimed invention of the present patent application. Kadansky et al. disclose a bottleneck-link speed, round-trip time, and hop count. Kadansky et al. also disclose static network with the maximum possible transmission rate, by the bottleneck bandwidth, for the static network in FIG. 10, the maximum possible transmission rate, as limited by the bottleneck bandwidth, is 50 (Kb/s). TRAM manages to keep the rate oscillating between 30 and 60. The initial spike is bigger, the result of slow start when there is no hint what the possible maximum rate is. The subsequent improved performance is what we expected. The fact that the cache occupancy oscillates at around 7 in the steady state is normal. In these tests, the ACK window is set to 8. With ACK staggering, the cache usage is expected to stay around 8, instead of going from zero (0) to eight (8) and back to zero (0) in a saw-tooth form. (col 33, 10-22), Round trip time (RTT) computations, (col 44, lines 45), every RxGroup-head computes the round trip time between itself and a RxGroup-member, adds its latency from the data source (learnt via its RxGroup-head) and relays the information to the RxGroup-member via the Hello-Uni message. Note that member latency reported in the Hello-Uni message refers to the previously computed latency. Until the RTT is computed, the Hello-Uni message, if generated, will have the receiver's member latency set to zero (0). Col 46, lines 7-15) and measuring by hop count. (col 16, lines 4-5). Kadansky et al. in col 15, lines 65-67 and col 16, lines 1-12 disclose that TRAM builds the tree so that repair heads are close to their members. A frame or message usually carries a "time-to-live" indication in its header, which was written by the source station, which transmitted the message onto the network. Upon expiration of the time-to-live value routers or other hardware of the network will no longer forward the message. The "time-to-live" is measured, in this document, by hop count. For example, if the time-to-live has the value of "2", the message will be forwarded for only two (2) hops.

Alternatively, if the time-to-live has a value of "4", the message will be forwarded over four (4) hops in the network. Use of a time-to-live designation in message headers

enables repair heads to perform repairs using small time-to-live (TTL) values, which not only minimizes network bandwidth consumption but also avoids unnecessary processing at receivers not requiring repair. By incorporate Gehr's idea of maximum server availability and load balancing with network related metric group within Kadansky's configuration in order to improve service, ensure high availability, enable load balancing and gracefully respond to faults, overloads and delays. Because the arguments with respect to the allowableness of independent claims were found unpersuasive, these same arguments are not persuasive with respect to the other dependent claims.

Claims 1-36 have been examined.

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

Claims 1-11, 13-23 and 25-35 are rejected under 35 U.S.C. 102(e) as being anticipated by Gehr et al. (US Patent 5,828,847)

1. As per claims 1,13 and 25, a method, an apparatus and a computer program product of balancing a workload across a plurality of servers, the method comprising the steps of:

responsive to a request from a requesting client for a distributed service, forwarding the request to a first distributed service manager associated with the requesting client;

determining whether the first distributed service manager has information about the distributed service; (corresponds to the dynamic server switching system maintaining a list in each client which identifies the primary server for that client, abstract)

if the first distributed service manager has information about the distributed service, retrieving the information about the distributed service, (corresponds to the event that the client does not have requests served by the designated primary server or the designated communication method, the system traverses the list to ascertain the identity of the first available alternate server-communication method pair. Abstract) and (this routing information functions to load balance on a per client basis since each client has its own routing list. Col 4, lines 22-25)

if the first distributed service manager does not have information about the distributed service, retrieving information about the distributed service from a second distributed service manager and caching the retrieved information at the first distributed service manager; and sending the retrieved information to the requesting client. (Client-Server System Philosophy, in the event that the client does not have requests served by the designated primary server or the designated communication method, then traverses the list to ascertain the identity of the first available alternative server-communication path pair, col 4,lines 58-66) and (see Fig 5.a and Fig 5.b)

2. As per claim 2, the method wherein the first distributed service manager has information about at least two sources for the distributed service and selects a source which will provide best service to the requesting client based on network performance metrics. (Load balancing is the practice of splitting communication into two or more routes. By balancing the traffic on each route, communication is made faster and more reliable. Load balance is not related to the absolute level of load, but only to how well the existing load is distributed. See also Fig 5.a and Fig 5.b)

3. As per claims 3, 15 and 27, a method, an apparatus and a computer program product of balancing demand for networked services in a distributed data processing system, the method comprising the steps of:

initializing one or more local service managers within the distributed data processing system, wherein each local service manager provides access to networked services for clients within the distributed data processing system, and wherein each client is uniquely associated with a local service manager, initializing one or more distributed service managers within the distributed data processing system, wherein each distributed service manager provides access to networked services to local service managers within the distributed data processing System, and wherein each local service manager is uniquely associated with a distributed service manager; receiving, at a distributed service manager, a request for a networked service from a local service manager; determining whether the distributed service manager has information about a networked service with one or more characteristics that match one or more parameters in the request for a networked service; and returning information about a matched networked service from the distributed service manager to the local service manager. (Corresponds to the dynamic server switching system maintaining a list in each client which identifies the primary server for that client, and to the event that the client does not have requests served by the designated primary server or the designated communication method, the system traverses the list to ascertain the identity of the first available alternate server-communication method pair. Abstract) and (Client-Server System Philosophy, col 4, lines 58-66) and (see Fig 5.a and Fig 5.b)

4. As per claim 4, the method further comprising: sending a request for a networked service from a requesting client to a local service manager associated with the requesting client; and

returning information about a matching networked service from the local service manager to the requesting client, wherein the matching networked service has characteristics that match parameters in the request for a networked service. (corresponds to the data generated as a result of the server actions are then returned to the requesting client. It should be noted that a server S1-S4 could be self-serving, in that it can also function as a client for some service requests, col 4, lines 1-11)

5. As per claim 5, the method further comprising: receiving a request for a networked service at a local service manager; and determining whether the local service manager has information about a networked service with characteristics that match parameters in the request for a networked service. (Corresponding entries are included in this data which designate preferred communication methods between client-server pairs, col 4, lines 18-20)

6. As per claim 6, the method further comprising:
if the local service manager has information about a matching networked service, returning the information about the matching networked service to the requesting client; if the local service manager does not have information about a matching networked service, forwarding the request for a networked service from the local service manager to a distributed service manager associated with the local service manager. (Client-Server System Philosophy, in the event that the client does not have requests served by the designated primary server or the designated communication method, then traverses the list to ascertain the identity of the first available alternative server-communication path pair, col 4, lines 58-66) and (see Fig 5.a and Fig 5.b)

7. As per claim 7, the method further comprising: if the distributed service manager has information about a matching networked service, returning the information about the matching networked service to the local service manager; if the distributed service manager does not have information about a matching networked service, broadcasting the request for a networked service from the distributed service manager to all distributed service managers in the distributed data processing system, receiving information about one or more matching networked services at the distributed service manager in response to the broadcast request; and caching the received information about one or more matching networked services at the distributed service manager. (Corresponds to the dynamic server switching system maintaining a list in each client which identifies the primary server for that client, Abstract) and (Corresponds to the event that the client does not have requests served by the designated primary server or the designated communication method, the system traverses the list to ascertain the

identity of the first available alternate server-communication method pair. Abstract) and (this routing information functions to load balance on a per client basis since each client has its own routing list. Col 4, lines 22-25) (Client-Server System Philosophy, in the event that the client does not have requests served by the designated primary server or the designated communication method, then traverses the list to ascertain the identity of the first available alternative server-communication path pair, col 4, lines 58-66) and (See Fig 5.a and Fig 5.b)

8. As per claim 8, the method further comprising: in response to a determination that the distributed service manager has information about two or more matching networked services, selecting a single networked service at the distributed service manager. (the client reestablishes the originally selected primary server-communication method pair as the request route for newly generated requests while processing existing requests via the alternate server which received these requests, col 4, line 67, and col 5, lines 1-5)

9. As per claim 9, the method further comprising:
performing a load balancing operation at the distributed service manager to select the single networked service. (the client reestablishes the originally selected primary server-communication method pair as the request route. Since each client has its own list, the load presented to each of the servers can be balanced across the available servers. Col 11, lines 38-41)

10. As per claim 10, the method further comprising:
comparing network-related metrics during the load balancing operation. (the client reestablishes the originally selected primary server-communication method pair as the request route for newly generated requests while processing existing requests via the alternate server which received these requests. (col 4, line 67, and col 5, lines 1-5)

11. As per claims 11, the method further comprising:
comparing one or more of network- related metrics associated with a network path between a requesting client and a providing server. (corresponds to the dynamic server

switching system avoids system bottlenecks and maintains a rapid exchange of communication between client and server so that the client processes obtain the benefit of the rapid data retrieval capability of the automated cartridge library system, col 11, lines 19-23)

12. Claims 14 and 26 recite the same limitations as claim 2. Therefore, they are analyzed and rejected by the same rationale.

13. Claims 16-23 and claims 28-35 recite the same limitations as claim 4-11. Therefore, they are analyzed and rejected by the same rationale.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 12, 24 and 36 are rejected under 35 U.S.C. 103(a) as being unpatentable over Gehr et al. (US Patent 5,828,847) and further in view of Kadansky et al. (US Patent No. 6,507,562)

14. As per claims 12, 24 and 36, Gehr et al. do not explicitly disclose the method wherein the network related metrics are selected from a group comprising: bottleneck-link speed, round-trip time, and hop count. However, Kadansky et al. Disclose static network with the maximum possible transmission rate, by the bottleneck bandwidth (col 33, 10-16), Round trip time (RTT) computations, col 44, lines 45 and measuring by hop count. (col 16, lines 4-5). Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention was made to incorporate Gehr's idea of maximum server availability and load balancing with network related metric group within Kadansky's configuration in order to improve service, ensure high

availability, enable load balancing and gracefully respond to faults, overloads and delays.

Conclusion

THIS ACTION IS MADE FINAL. Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Mitra Kianersi
Feb/07/2005

JASON CARDONE
PRIMARY EX
AU 2145